DOCUMENT RESUME

ED 448 403 CG 030 625

AUTHOR Aviles, Christopher B.
TITLE Grading with Norm-Referenced or Criterion-Referenced Measurement: To Curve or Not To Curve, That Is the Question.
PUB DATE 2001-00-00
NOTE 15p.
PUB TYPE Opinion Papers (120) -- Reports - Descriptive (141)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS College Faculty; College Students; *Criterion Referenced Tests; Grading; Higher Education; *Measurement Techniques; *Norm Referenced Tests; Required Courses; *Social Work; *Student Evaluation; *Test Interpretation

ABSTRACT

Assigning grades is an integral part of social work education. However, social work educators must decide whether to use norm-referenced or criterion-referenced measurements to grade exams and other assignments. This paper presents arguments for grading with both norm-referenced and criterion-referenced measurements. The benefits of criterion-referenced measurement as a choice for one professor's classes in social work education are reviewed. One criticism of this measure questions whether grades are devalued when many others attain the same achievement. The conclusion is made that it is not possible to examine professors' grade spreads in order to learn anything about their instructional decisions, techniques, and testing that generated these grades. New professors should not be examined based on the highs and lows of their exams as an indication of the standards in their classes. But, by using criterion-referenced measurements, instructors can compare student achievement to their chosen standard instead of to the achievement of other students. (JDM)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 448 403

TITLE

Grading with norm-referenced or criterion-referenced measurement:
To curve or not to curve, that is the question.

AUTHOR

Christopher B. Aviles, Ph.D., ACSW
Assistant Professor of Social Work

Social Work Department
Buffalo State College
1300 Elmwood Avenue
Buffalo, NY 14222

Work 716-878-5327

avilescb@buffalostate.edu

BEST COPY AVAILABLE

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☐ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
C AVILES
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

CG030625

2

Abstract

Assigning grades is an integral and everyday part of social work education. However, social work educators must decide whether to use norm-referenced or criterion-referenced measurement to grade exams and other assignments. Norm-referenced measurement is commonly called grading on a curve in academia. I was not clear about the difference between the two types of grading as a new social work educator 12 years ago. Many exams and papers later, I am clear about the difference. While grading on the curve is not dead in academia, I have eliminated all traces of it in my courses. New social work educators and, perhaps, veteran social work educators may benefit from a review of both types of grading.

This paper examines both sides of a common grading controversy. Grading with norm-referenced and criterion-referenced measurement are reviewed along with issues related to both types of grading. I will describe why I grade with criterion-referenced measurement and believe it is a better choice for social work education.

THE ISSUE

"Professor, I scored the highest in the class with 60% of 100%. What grade is that?"

Grading on "a curve" has long been an accepted practice in academia. Amidst talk of increasing academic standards and measuring student outcomes, it is time to challenge the practice of grading on the curve and have social work educators think more deliberately about grading. As a new social work educator 12 years ago, I had questions and doubts about grading my first exam that other new social work educators may have. "How do I tell the difference between a grade of A and a grade of B? How many students will (and should) excel or fail? What do my grades say about me as a new instructor?" I also received advice (and warnings) from senior faculty about what grades say about an educator. For example, a senior instructor toured me around our building in my first semester in order to view midterm exam grades posted outside the classrooms. He explained that instructors with many A grades were "easy instructors with low standards" (a bad thing) and instructors who assigned many failing grades were "good instructors with high standards" (a good thing). I recall making a mental note: all students flunk = excellent instructor. Although instructors are free to decide how to grade, grades can be interpreted differently by colleagues when exam score distributions do (or do not) deviate from normal.

Measuring outcomes, raising standards, and increasing student achievement are serious issues getting much attention lately. However, I challenge social work educators to consider the practical and often difficult task

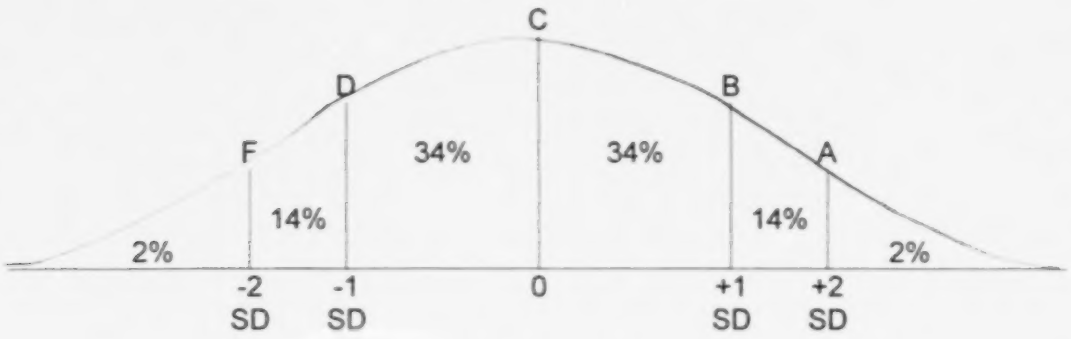
of grading exams. This article is intended to encourage new social work educators to think deliberately about grading and to challenge veteran social work educators to rethink grading on a curve.

Assigning grades, or more properly, measuring student achievement, is normally done with either norm-referenced or criterion-referenced measurement and social work educators must choose between them. Let's define both approaches for the new social work educators. For illustration, the grading examples will assume that exam scores are generated from a 100-question objective format exam where each question is worth one point. The exam generates a score that is reported as percent correct of 100% (ex: 85% correct of 100%), or reported as a raw score of the number of questions correct of 100 questions (ex: 85 answered correctly of 100 questions).

Norm-referenced Measurement

The purpose of grading with norm-referenced measurement is to separate students' based on achievement level by comparing their achievement to the achievement of other students (Gentile, 1990). Norm-referenced measurement is ordinarily called grading on the "curve" because a normal distribution of scores, or bell curve, results despite the range of exam scores (Figure 1). Norm-referenced measurement is useful when students must be ranked for something with a limited number of spaces, e.g., for college admission or awarding scholarships.

Fig. 1. Norm-referenced letter grades from standard deviations



Social work educators who grade with norm-referenced measurement simply calculate a class mean exam score and assign letter grades based on the standard deviations. Campus test scoring services routinely provide instructors with these descriptive statistics. Figure 1 highlights the relationship between numerical exam scores and norm-referenced letter grades (Note: the curves are drawn for illustration and are not perfect). Fifty percent of any class scores above and below whatever the median exam score is and students score one or two standard deviations above and below whatever the mean exam score is. Normally the highest exam score receives a grade of A and the lowest score a grade of F regardless of the actual exam score. For example, if the highest class exam score is 60% of 100%, the score is two standard deviations above the mean score and is a letter grade of A. Alternatively, if 90% of 100% is the lowest score, it is two standard deviations below the mean score of 95% and is a grade of F. It is common to post exam scores, ordered from highest to lowest, outside classrooms, with lines drawn at the cutoff scores for each letter grade. There are no rules for assigning letter grades and a social work educator can simply decide

that two standard deviations above the mean score is a grade of B instead of an A.

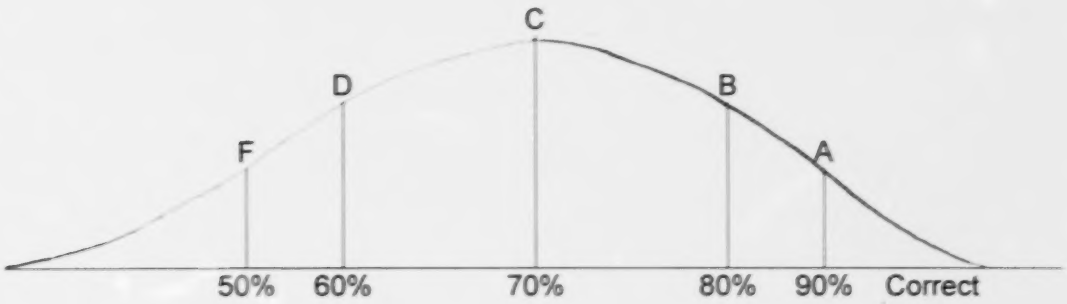
My students sometimes say they have had instructors in other academic departments who announce in the first class meeting that there will be X amount of A grades in the class. These instructors have probably decided that the two percent of exam scores that fall two standard deviations above the mean will get a grade of A (despite the actual exam score). Assuming an instructor always has 100 students per class, they know on the first class day (and for the rest of their academic careers) that 2% of the class or two students will get a grade of A. Norm-referenced grading is also easily applied to written projects. The best X papers (based on class size) get a grade of A and the worst X papers get an F.

Criterion-referenced Measurement

Criterion-referenced measurement compares student achievement to an instructor chosen standard instead of to the achievement of other students. If an instructor decides an exam score of 90% of 100% is the criterion or standard for a letter grade of A, all students scoring 90% or better get an A. If the highest-class exam score is 80%, no one gets an A (Figure 2). Social work educators who grade with criterion-referenced measurement use cutoffs for letter grades based on instructor chosen standards (commonly percents) instead of with standard deviations. Traditionally, the following cutoffs often correspond to letter grades: A = 90% -100%, B = 80%-89%, etc. An instructor can choose a different percentage and perhaps make 95% the standard for a grade of A. Criterion-referenced measurement may produce "abnormal or skewed" score distributions

because all students can statistically meet (or not meet) the criterion (Gronlund, 1981; Martuza, 1977).

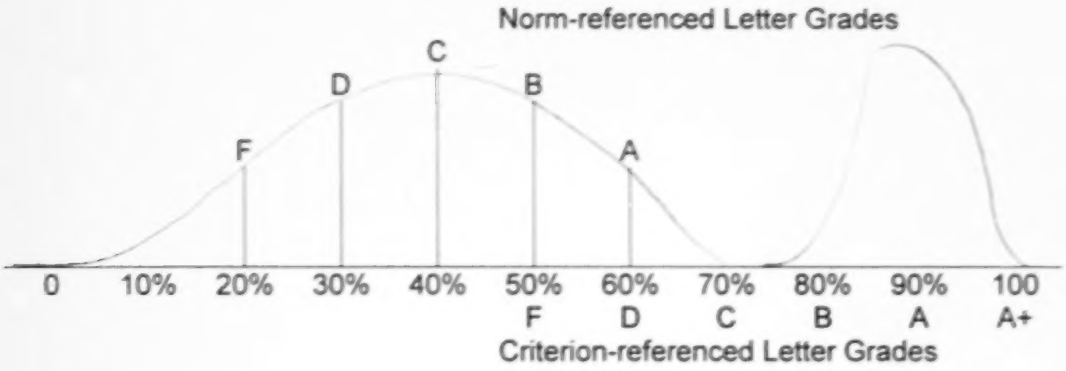
Fig. 2. Criterion-referenced letter grades from percent correct of 100%



The teaching method called mastery learning utilizes criterion-referenced grading and proponents predict it will produce achievement gains of two standard deviations (Bloom, 1977). The claims are statistically possible with criterion-referenced measurement. This means 90% of students can score in the range statistically reserved for the top 10%. Said differently, an entire class earns an A when the lowest class exam score is 90%. In contrast, with norm-referenced measurement, 90% converts to a grade of F because it is the lowest class score. With criterion-referenced grading, an entire class gets a D if the highest exam score is 60%.

Figure 3 compares letter grades generated from both norm- and criterion-referenced measurement. Assuming an exam score of 60% is the highest-class score, it is a letter grade of A with norm-referenced measurement and a grade of D with criterion-referenced measurement.

Fig. 3. Norm-referenced and criterion-referenced letter grades



Said simply, norm-referenced measurement helps social work educators determine which students achieve the highest when compared to other students. Criterion-referenced measurement helps social work educators determine whether students achieve to the levels we expect from them.

ONE SOCIAL WORK EDUCATOR'S CHOICE

As an undergraduate social work educator, I prefer criterion-referenced grading for several reasons. I have serious reservations about saying all the material I teach is important and then potentially giving an A grade to students who only score 60% of 100% on a test of that "important material" (assuming 60% = highest class score). How do I know what 40% of the "important material" students lacked and what 60% they had? The professors who teach the second part of multi part courses often know (No, professor, we did not get that far in Human Behavior 1; No, professor, we never learned that.)

I am also concerned that grading on a curve may mask my poor teaching, since a normal score distribution results regardless of what I do in the classroom.

Grading on a curve makes it difficult to measure if teaching skill has improved (No matter what I do to improve my teaching skill, each semester 50% of my students score below the median and only 2% get an A!). If grading on a curve can mask what happens in a classroom, criterion-referenced grading does the reverse by forcing a social work educator to ask "what happened" when the highest class score is 60%. I warn my social work students to avoid the "rookie" mistake of always interpreting client success as a positive statement about the SOCIAL WORKER and client failure as a statement about the CLIENT'S unwillingness to engage in intervention. The same caution applies to new social work educators (and perhaps veterans also) who use criterion-referenced grading and have student achievement below what is expected. In this case, you may have to ask whether your expectations were too high or the effort of students was too low.

I have never compared an exam score of one student (say, 76%) to another student (say, 82%) and made some instructional decision based on the comparison. I regularly compare a student's score (say, 89%) to what I expect them to score on an exam and use traditional percent cutoffs to assign a letter grade (89% = B). I am less concerned about where student X falls compared to student Z and more concerned about where both fall compared to my learning expectations. I am concerned that norm-referenced grading may not prepare my students for those graduate schools where students perform against standards and not against other students. In certain situations, like deciding on admissions

to departments or schools with limited space, it makes sense to use norm-referenced measurement to compare students, but not in the classroom.

Student Reactions

Students appear aware of norm- and criterion-referenced measurement but they do not use these terms. I use the following sports analogy when my class asks if I "curve." "First place in an Olympic race wins the gold medal even if the race time was the slowest in Olympic history. That's grading on a curve. Criterion-referenced grading means you must set a new Olympic record for the gold medal and not just beat the other racers." Students often call this "straight cutoffs," probably meaning that 90% of 100% correct is a grade of A, 80-89% = B, etc. Students often have one of two reactions to criterion-referenced grading. Some appear relieved they will not be competing against classmates for a limited number of grades. Other students appear unable to gauge their achievement without comparing it to their classmates. For example, after scoring high on an exam some of my students say they believed they learned much of the material, but were disappointed because so many other students also earned an A grade ("I guess I did not learn as much as I thought."). At the other extreme, one student apparently forgot that I do not "curve" and exclaimed after finding he scored the highest on a test my entire class failed: "I'm number one!"

FINAL THOUGHTS

I have seen instructors advocate, often strenuously, for one of the other type of grading and noted much emotion associated with both. For example,

norm-referenced and criterion-referenced "graders" can both claim the other produces devalued grades but for different reasons. Criterion-referenced graders can say grades produced from norm-referenced measurement are devalued because they occur regardless of the exam scores. Earning the highest class grade may not be a great achievement if the score is 40% of 100%. I would not want my oral surgeon scoring the highest in his/her graduating class with 40% of 100% (Hopefully he/she passed the novocaine class!).

Norm-referenced graders can say grades produced from criterion-referenced measurement are devalued when more than expected occur because achievement is devalued when others attain the same achievement. Thus, a grade of A is more valuable when fewer occur. Grades, therefore, become a commodity, rising and falling in worth based on scarcity. However, does scarcity equate with achievement? Said differently, are fewer A grades and more failing grades always the result of increased standards? As I learned on my "rookie tour of the building" mentioned earlier, some educators may believe so. It was perhaps in this spirit that while serving on a committee charged with finding ways to increase campus standards, an instructor offered us a simple three word plan to raise standards: fail more students. This plan assumes that increased failure is the result of increased standards and not low quality instruction.

One might say that proponents of both "camps" draw battle lines in the sand and take new recruits on patrol in the halls of their buildings to find grade spreads. Norm-referenced graders who find a class with many A grades can say, "This instructor has low standards and easy tests!" Criterion-referenced graders

upon finding a class where an A grade is an exam score of 60% can say, "This instructor has high standards, but doesn't require their students to meet them!"

In reality, it is not possible to examine grade spreads and know anything about the instructional decisions, techniques and testing that generated them. Colleagues can still say (and have said to me) someone is an easy instructor with low standards because many students (more than two percent) earned grades of A. However, in 12 years of teaching no one has ever (and I mean never) asked me for the difficulty index statistic on any exam item or for an entire exam. No one has ever asked if my exam tested the lower levels of Bloom's (1956) taxonomy of educational objectives (knowledge, comprehension) or tested the higher levels that constitute critical thinking (application, synthesis, analysis, evaluation). No one has ever asked if my tests employ near transfer of knowledge (at worst, repeating what was taught in class) or far transfer (applying principles to unique situations students may encounter in the field). No one has ever asked if I used my own exams or exams created by colleagues, graduate students, or textbook publishers. New social work educators should be aware that others might examine your grade spreads and "see" low or high standards and hard or easy exams.

I hope I have challenged some of you to abandon grading on the curve. I also hope this article helps new social work educators decide what grading method to employ, instead of using whatever the "grading method du jour" is in your department, or worse, grading as you were graded as a student. Who knows how our own teachers chose the grading methods they did.

Let's close with a question that new social work educators will no doubt have to answer early in their careers: Professor, I scored the highest in the class with a 60% of 100%. What grade is that?

REFERENCES

Bloom, B., S. (1956). *Taxonomy of educational objectives: Cognitive domain.* (NY: Longman).

Bloom, B., S. (1977). Time and learning. In M. Wittrock (Ed.), *Learning and Instruction* (pp. 586-597). (Berkeley, CA: McCutchan).

Gentile, J., R. (1990). *Educational psychology.* (Dubuque, IO: Kendall\Hunt).

Gronlund, N., G. (1981). *Measurement and evaluation in teaching* (4th ed.). NY: Macmillan.

Harris, C., S. (1974). Some technical characteristics of mastery tests. In C. Harris H, M. Alkin C, & W. Popham J (Eds.), *Problems in criterion-referenced measurement* (pp. 98-115). Los Angeles, CA: Center for the Study of Evaluation.

Martuza, V., R. (1977). *Applying norm-referenced and criterion-referenced measurement in education.* (Boston, MS: Allyn and Bacon).